DOCUMENT RESUME

ED 340 734 TM 017 788

AUTHOR Halpin, Gerald; And Others
TITLE So You Have Chosen an Unequal Cell Size ANOVA Option--Do You Really Know What You Have?
PUB DATE Nov 91
NOTE 27p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (20th, Lexington, KY, November 12-15, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Computer Software Evaluation; *Decision Making; Equations (Mathematics); Mathematical Models; *Regression (Statistics); *Research Design; *Sample Size
IDENTIFIERS *Nonorthogonal Variables; Statistical Analysis System; Statistical Package for the Social Sciences; *Variance Partitioning

ABSTRACT

The objective of this investigation was to take a small data set that represents unbalanced factorial designs and explain by example how the variance is partitioned when using the various options from the Statistical Package for the Social Sciences (SPSSX) and Statistical Analysis System (SAS). That the unequal cell size analysis of variance (ANOVA) is in the typical situation a special case of multiple regression is demonstrated. Specifically, the study describes how the variance is being partitioned when options 9 (unique) or 10 (hierarchical), or default from SPSSX and Type I or Type III sums of squares options from SAS are chosen. Data (N=39 scores) used in the demonstration analyses using the different methods have three levels of Factor A and two levels of Factor B, and the number of observations in the cells are not equal. The analytic examples give researchers a better idea of what is happening when different sums of squares options in SAS or various options in SPSS are used. Two tables present data from the analysis and five figures illustrate the partitioning. A 10-item list of references is included. (SLD)

ED340734

So You Have Chosen An Unequal Cell Size ANOVA Option--

Do You Really Know What You Have?

Gerald Halpin, Regina Carwile, and Glennelle Halpin

Auburn University

Paper presented at the meeting of the Mid-South Educational Research Association, Lexington, November, 1991

Abstract

The objective of this investigation was to take a small data set which represents the unbalanced designs and explain by example how the va.·iance was actually partitioned when utilizing the various options from the SPSS and SAS statistical packages. That the unequal cell size ANOVA is in the typical situation a special case of multiple regression will be demonstrated. The analytic examples provided will give researchers a better understanding of what is happening when different sum of squares options in SAS or various options in SPSS are employed.

Analysis of variance (ANOVA) is arguably the most widely utilized statistical procedure in education and the social sciences (Edington, 1974; Goodwin & Goodwin, 1985; Halpin & Halpin, 1988; Willson, 1980). It might further be argued that unbalanced factorial designs are the most widely employed ANOVA designs. The authors believe that it is the rule rather than the exception for researchers to have unequal cell sizes in their investigations. We tend to be skeptical of the investigations where researchers report equal cell sizes, especially when they fail to explain the procedures utilized to obtain equal cell sizes. This skepticism remains regardless of whether or not the research design is experimental or correlational in nature.

In experimental research investigators have unbalanced designs for many reasons. Subjects miss treatment and testing sessions due to such things as sickness or conflicting activities. They may withdraw from the experiment or at times be uncooperative and refuse to respond to treatments and/or questions on response measures. Equipment breakdown is not uncommon, and experimenters make errors. Thus, missing data and unequal cell sizes exist even in the most competently planned research.

In nonexperimental research, unequal cell sizes usually reflect reality when there are naturally occurring variables such as race, socioeconomic status, and religious affiliation. Taking steps to create equal cell sizes under such conditions regardless of the temptation creates more problems than solutions. Even though the only completely acceptable solution to the missing data problem is to not have any (Cochran & Cox, 1950), we cannot and should not throw our data away simply because we no longer have equal sample sizes.

Assume that you had two independent variables, race and socioeconomic status, in a two-factor ANOVA problem. Given what is known about the relationship between race and socioeconomic status, it is almost certain that substantial differences will exist in cell sizes. Substantially more blacks are going to be in the lower socioeconomic status group and more whites in the upper socioeconomic group. When this problem is encountered, a typical solution is to drop subjects until cell sizes are equal. The establishment of equal cell sizes when in fact the cell sizes are not equal in reality results in what Humphreys and Fleishman (1974) refer to as "pseudo-orthogonal" designs and results in what Hoffman (1960) refers to as the "dismemberment of reality." Making naturalistically occurring variables such as race and socioeconomic status independent by dropping cases in the factorial ANOVA would generate unrealistic results. Therefore, under most conditions unequal size factorial ANOVA is either unavoidable or desirable.

The ANOVA procedure in the Statistical Package for the Social Sciences--SPSSX (Norusis, 1988) or SPSS/PC+ (Norusis, 1990)--or the General Linear Model procedure from the Statistical Analysis System--SAS (SAS Inc., 1985)--are typically employed when dealing with the unequal cell sizes. What are the differences between these approaches? Do researchers who make these choices really know how the variance is being partitioned?

The purpose of this undertaking is to explain how the variance is actually being partitioned using the various options from the SPSS and SAS statistical packages when the researcher has unequal cell sizes. More specifically, our objective is to describe how the variance is being partitioned when Option 9 (unique), Option 10 (hierarchical), or default from SPSSX or SPSS/PC+ and Type I or Type III sums of squares options from SAS are chosen. When researchers understand how the variance is actually being partitioned, they will be better able to match the appropriate analytical option with their research questions.

Method

In the classical factorial analysis of variance model the total variance or sum of squares is partitioned into mutually exclusive components reflecting various effects. In the two-factor completely randomized design the total sum of squares is partitioned into the sum of squares for Factor A, the sum of squares for Factor B, the sum of squares for the interaction of Factor A and Factor B, and the sum of squares for the residual or error as reflected in Figure 1.

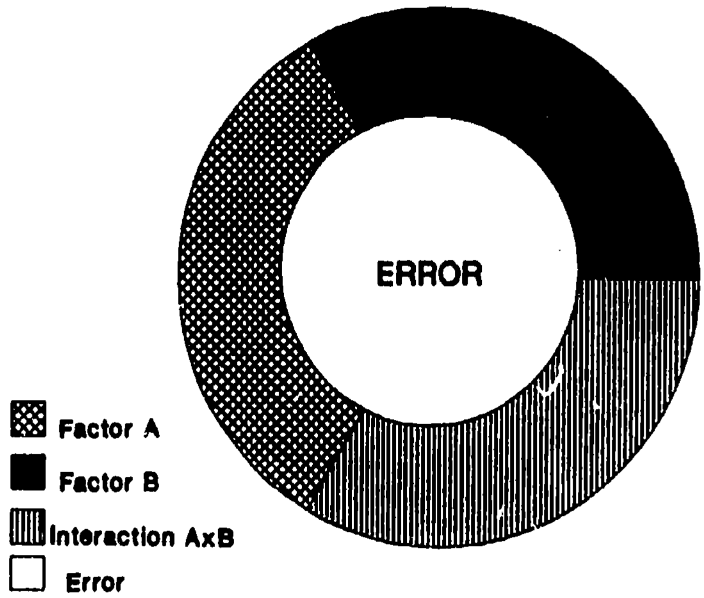

Figure 1. Variance partitioned for equal cell sizes.

These various effects are unambiguously partitioned into mutually exclusive and exhaustive categories as presented in Figure 1 when the cell sizes of the design are equal.

However, problems occur when for any reason the cell sizes become unequal. Unequal cell size designs are frequently referred to as unbalanced or nonorthogonal designs since Factor A, Factor B, and the interaction between Factors A and B are intercorrelated. In multiple regression terminology we have a multicolinearity problem. The total sum of squares no longer equals the sum of squares of Factors A, B, A x B interaction, and error.

In the two-factor case the effects of A, B, and the A x B interaction not only are interrelated but also share in the accounting of the variance of the dependent variable. In Figure

2, the different shaded areas represent the unique effects of Factors A, B, and the A x B interaction. The shaded areas labeled 1, 2, and 3 separating each of the unique effects represent the proportions of the dependent variable variance accounted for jointly by the effects on either side.

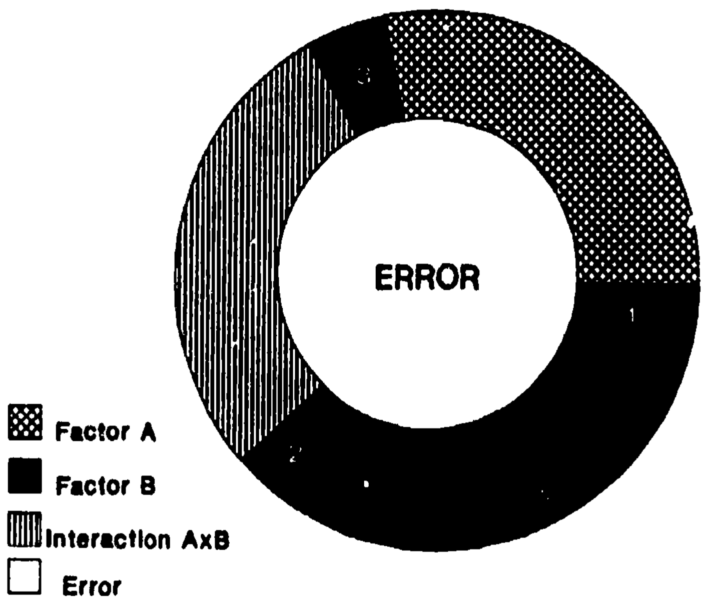

Figure 2. Unequal cell size analysis of variance.

Because the dependent variable variance can no longer be unambiguously partitioned among the different main effects and the interaction, researchers may ask how the variance is to be assigned to the different effects. Applebaum and Cramer (1974) observed that "[t]he nonorthogonal multifactor analysis of variance is perhaps the most misunderstood analytic technique" and we might add one of the most controversial techniques "available to the behavioral scientists, save factor analysis: (p. 335).

The three most frequently used methods for partitioning the variance in the nonorthogonal factorial designs were explained by Overall and Spiegel (1969). In the two-factor design their Method 1, Method 2, and Method 3 correspond respectively to Option 9 (unique), default option, and Option 10 (hierarchical) in the ANOVA procedure from SPSS. In the GLM procedure from the SAS system, the Type I sums of squares option is equivalent to SPSS Option 10 and Overall and Spiegel's Method 3, and Type III sums of squares from SAS is equivalent to Option 9 of SPSS and Method 1 from Overall and Spiegel. The interested reader would profit from reading Overall and Spiegel (1969) and Lutz (1979).

The various methods of dealing with the nonorthogonal ANOVA problem via SPSS and SAS can probably best be understood as a multicolinearity multiple regression problem. First, we need to restate that with the two-factor orthogonal ANOVA design the total sum of squares is equal to the sum of the sum of squares for Factor A, Factor B, the interaction of A x B, and error as presented in Figure 1. From a multiple regression perspective the sum of squares for each effect and error can be divided by the total sum of squares to yield the proportions of variance accounted for by each source. These proportions are known as R^2s, and the sum of these proportions of variance for all of the sources (PV^2_T) equals 1. Equation 1 reflects the two-factor orthogonal case.

$$\text{Equation 1.} \quad PV^2_T = R^2_A + R^2_B + R^2_{AxB} + (1 - R^2_{MAX})$$

In Equation 1, the total proportion of variance (PV^2_T) is equal to the proportion of variance accounted for by Factor A (R^2_A) plus the proportion of variance accounted for by Factor B (R^2_B) plus the proportion of variance accounted for by the A x B interaction (R^2_{AxB}) plus the proportion of error variance ($1 - R^2_{MAX}$). As can be observed in Figure 1, the sums of the areas within the circle would equal 1 if those areas are converted to proportions as is being discussed here.

The dilemma of jointly accounting for the variance in the dependent variable occurs when unequal cell sizes exist. This problem can be observed in Figure 2 where the total proportion of variance (PV^2_T) is no longer equal to the proportion of variance accounted for by Factor A (R^2_A) plus the proportion of variance accounted for by Factor B (R^2_B) plus the proportion of variance for interaction (R^2_{AxB}) plus the proportion of variance for error ($1 - R^2_{MAX}$) as reflected in Equation 2.

$$\text{Equation 2.} \quad PV^2_T \neq R^2_A + R^2_B + R^2_{AxB} + (1 - R^2_{MAX})$$

The proportions of variance for Factors A, B, and the A x B interaction are represented as in Figure 1 except that in Figure 2 there are wide shaded boundaries labeled 1, 2, and 3 between each of the effects. These overlapping areas in Figure 2 represent the proportion of the total variance of the dependent variable jointly accounted for by Factors A and B (Area 1), Factor B and the A x B interaction (Area 2), Factor A and the A x B interaction (Area 3). If these areas of overlap are allocated to Factor A, Factor B, and the interaction of A x B, then the variance will be accounted for twice and the total proportion of variance explained will be greater than 1. Stated differently, the sums of squares for the various effects plus error will be greater than the total sum of squares. If the overlap areas are not assigned to one of the effects, the total proportion of variance allocated will be less than 1, and the sum of squares for the various effects plus the sum of squares for error will be less than the total sum of squares. This sharing of dependent variable variance by more than one independent variable represents the widely known multicolinearity multiple regression problem, and the different procedures for dealing with this problem are central in this paper.

Data Set

Before considering the explanations of the methods, observe in Table 1 the data employed in demonstration analyses using the

Insert Table 1 about here

different methods. Note that there are three levels of Factor A and two levels of Factor B and the number of observations in the cells are not equal. Next, peruse Table 2 and observe that the

- - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - -

sums of squares, F ratios, probability levels, and proportions of variance (R^2) by each of the effects differ for all three methods. These differences among the three methods are small and of little consequence with the data set we have chosen to analyze. If the discrepancy among the cell sizes had been greater the multicolinearity problem would have been greater and the potential for disparate outcomes among the three methods would have increased.

SPSS Default Method

When cell sizes are unequal and the researcher fails to specify an option, the default option is utilized with SPSS. Type I and Type III sums of squares are routinely provided in SAS, neither of which compares with the default option of SPSS.

Utilizing multiple regression concepts, we explain how the sums of squares, F ratios, and proportions of variance in Table 2 are obtained for the SPSS default option. In explaining the results we will utilize Equation 3 and Figure 3.

Equation 3. $PV^2_T = R^2_{A.B} + R^2_{B.A} + R^2_{AxB.A,B} + R^2_{O1} + (1 - R^2_{MAX})$

In utilizing Equation 3 to explain the default ANOVA results in Table 2, the proportion of total variance in the dependent variable (PV^2_T) is equal to the proportion of variance accounted for by Factor A while controlling for Factor B ($R^2_{A.B}$) plus the proportion of variance accounted for by Factor B while controlling for Factor A ($R^2_{B.A}$) plus the proportion of variance accounted for uniquely by the A x B interaction while controlling for the main effects of both Factors A and B ($R^2_{AxB.A,B}$) plus the proportion of variance accounted for by both Factors A and B but not attributed to either (R^2_{O1}) plus the error variance which is that proportion of dependent variable variance not accounted for by any of the four specified effects ($1 - R^2_{MAX}$). The unique aspect of the default option method of analysis from SPSS as depicted in Equation 3 is R^2_{O1}. R^2_{O1} as well as the other aspects of the default method of analysis can probably best be depicted by utilizing the information in Table 2 and referring to Figure 3.

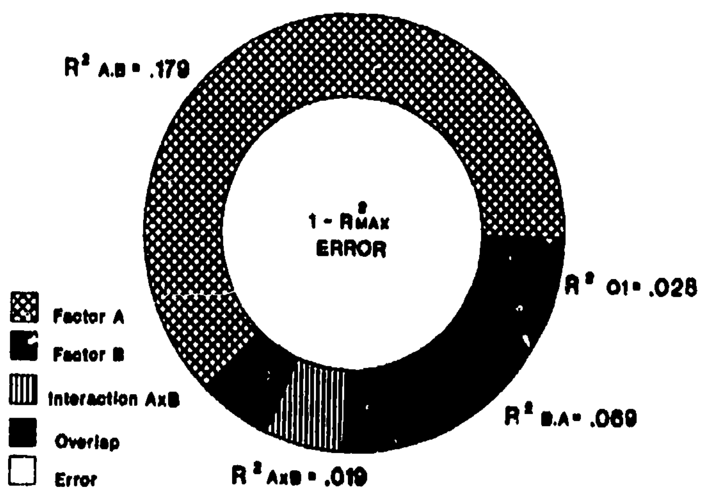

Figure 3. Variance partitioned using default option.

Referring to Table 2 under the default method, we find that the proportion of variance allocated to Factor A is obtained by dividing the sum of squares for Factor A while controlling for Factor B ($SS_{A.B}$) by the total sum of squares (SS_{TOTAL}). $R^2_{A.B} = SS_{A.B} / SS_{TOTAL} = 6.806/38.124 = .179$ with an F ratio = 4.307 and a probability of .022. In Figure 3, $R^2_{A.B} = .179$ depicts the proportion of dependent variable variance uniquely represented by Factor A plus the overlap area between Factor A and the A x B interaction labeled 3.

The proportion of variance allocated to Factor B is found by dividing the sum of squares for Factor B controlling for Factor A ($SS_{B.A}$) by the total sum of squares (SS_{TOTAL}). $R^2_{B.A} = SS_{B.A} / SS_{TOTAL} = 2.617/38.124 = .069$ with an F ratio = 3.312 and a probability of .078. In Figure 3, $R^2_{B.A} = .069$ represents the proportion of the dependent variable variance accounted for uniquely by Factor B plus the area of overlap labeled 2 which is the proportion of the dependent variable variance accounted for jointly by Factor B and the A x B interaction.

The proportion of variance allocated to the interaction is obtained by dividing the sum of squares for interaction controlling for Factors A and B ($SS_{AxB.A,B}$) by the total sum of squares (SS_{TOTAL}). $R^2_{AxB.A,B} = SS_{AxB.A,B} / SS_{TOTAL} = .731/38.124 = .019$ with an F ratio = .463 and a probability of .634. In Figure 3, $R^2_{AxB.A,B} = .019$ represents the proportion of dependent variable variance accounted for uniquely by the A x B interaction after controlling for Factors A and B.

But not yet accounted is the overlap area labeled 1. This portion of variance in the dependent variable is referred to as R^2_{01} in Equation 3. The area of overlap labeled 1 (R^2_{01}) in Figure 3 represents the dependent variable variance accounted for by both Factors A and B and not assigned to any effect: $R^2_{01} = 1 - [R^2_{A.B} + R^2_{B.A} + R^2_{AxB.A,B} + (1 - R^2_{MAX})] = 1 - [.179 + .069 + .019 + .705] = .028$.

The area labeled ERROR in Figure 3 refers to the proportion of variance in the dependent variable not allocated to either main effect, the A x B interaction, or R^2_{01}. In Table 2, the proportion of the dependent variable variance designated as error is found by dividing the sum of squares for error by the total sum of squares: $(1 - R^2_{MAX}) = SS_{ERROR} / SS_{TOTAL} = 26.866/38.124 = .705$.

SPSS Option 9 and SAS Type III Sums of Squares

As with the default method the multiple regression concepts are utilized to explain how the sums of squares, F ratios, and proportions of variance in Table 2 are obtained for the SPSS Option 9 and SAS Type III sums of squares. In explaining the results we will utilize Equation 4 and Figure 4.

Equation 4. $$PV^2_T = R^2_{A.B,AxB} + R^2_{B.A,AxB} + R^2_{AxB.A,B} + R^2_{02} + (1 - R^2_{MAX})$$

The proportion of total variance in the dependent variable (PV^2_T) is equal to the proportion of variance accounted for by

Factor A while controlling for Factor B and the A x B interaction ($R^2_{A.B,AxB}$) plus the proportion of variance accounted for by Factor B while controlling for Factor A and the A x B interaction ($R^2_{B.A,AxB}$) plus the proportion of variance accounted for by the A x B interaction while controlling for the main effects of both Factors A and B ($R^2_{AxB.A,B}$). The proportions of variance allocated to the main effects and interaction effect will not include any of the overlap areas. This jointly accounted for variance in the dependent variable (R^2_{O2}) is equal to the proportion of variance accounted for jointly by Factors A and B (overlap labeled 1 in Figure 4) but not attributed to either plus the proportion of variance accounted for jointly by Factor A and the A x B interaction (overlap labeled 3 in Figure 4) but not attributed to either plus the proportion of variance accounted for jointly by Factor B and the A x B interaction (overlap labeled 2 in Figure 4) but not attributed to either plus the error variance which is not accounted for by any of the four specified effects. The unique aspect of the Option 9 ANOVA procedure from SPSS and Type III sums of squares from the GLM procedure of SAS in Equation 4 is R^2_{O2}, the proportion of variance in the dependent variable which is left unallocated to either main effect, the interaction effect, or error. R^2_{O2} along with the other aspects of SPSS Option 9 and SAS Type III sums of squares methods of analyses can probably best be depicted by utilizing the information in Table 2 and referring to Figure 4.

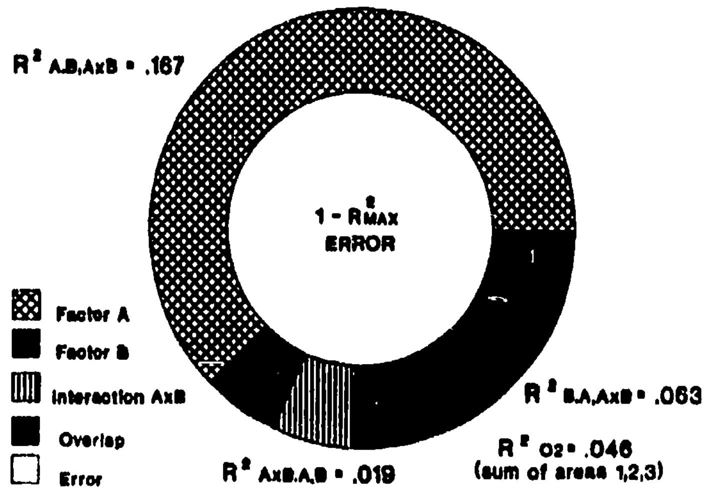

Figure 4. Variance partitioned using SPSS Option 9 and SAS Type III sums of squares.

Referring to Table 2 under the Option 9/Type III method, we find that the proportion of variance allocated to Factor A is found by dividing the sum of squares for Factor A while controlling for Factor B and the A x B interaction ($SS_{A.B,AxB}$) by the total sum of squares (SS_{TOTAL}). $R^2_{A.B,AxB} = SS_{A.B,AxB} / SS_{TOTAL} = 6.271/38.124 = .167$ with an F ratio = 4.032 and a probability of .027. In Figure 4, $R^2_{A.B,AxB} = .167$ represents the proportion of dependent variable variance accounted for by Factor A while controlling for Factor B and the A x B interaction.

Similarly, the proportion of variance allocated to Factor B is found by dividing the sum of squares for Factor B while controlling for Factor A and the A x B interaction ($SS_{B.A,AxB}$) by the total sum of squares (SS_{TOTAL}). $R^2_{B.A,AxB} = SS_{B.A,AxB} / SS_{TOTAL} = 2.391/38.124 = .063$ with an F ratio = 3.026 and a probability of .091. In Figure 4, $R^2_{B.A,AxB} = .063$ represents the proportion of dependent variable variance accounted for by Factor B while controlling for Factor A and the A x B interaction.

The SPSS Option 9/SAS Type III sum of squares methods to determine the proportion of dependent variable variance accounted for uniquely by the A x B interaction are identical to the default approach. The proportion of variance was obtained by dividing the sum of squares for interaction ($SS_{AxB.A,B}$) by the total sum of squares (SS_{TOTAL}). $R^2_{AxB.A,B} = SS_{AxB.A,B} / SS_{TOTAL} = .731/38.124 = .019$ with an F ratio = .463 and a probability of .634. In Figure 4, $R^2_{AxB.A,B} = .019$ represents the proportion of dependent variable variance accounted for by the A x B interaction while controlling for Factors A and B. Recall from the default option that none of the areas of overlap labeled 1, 2, and 3 are allocated to the A x B interaction effect.

The area labeled R^2_{O2} in Figure 4 represents the proportion of variance accounted for jointly by Factors A and B and not assigned to any effect (area of overlap labeled 1) plus the proportion of variance accounted for jointly by Factor B and the A x B interaction (area of overlap labeled 2) plus the proportion of variance accounted for jointly by Factor A and the A x B interaction (area of overlap labeled 3). $R^2_{O2} = 1 - [R^2_{A.B,AxB} + R^2_{B.A,AxB} + R^2_{AxB.A,B} + (1 - R^2_{MAX})] = 1 - [.167 + .063 + .019 + .705] = .046$.

The area labeled ERROR in Figure 4 refers to the proportion of variance in the dependent variable not allocated to either main effect, the A x B interaction, or R^2_{O2} (overlaps labeled 1, 2, 3). In Table 2 the proportion of dependent variable variance designated as error is found by dividing the sum of squares for error by the total sum of squares: $(1 - R^2_{MAX}) = SS_{ERROR} / SS_{TOTAL} = 26.866/38.124 = .705$. The proportion of dependent variable variance labeled error is the same as in the default option.

SPSS Option 10 and SAS Type I Sums of Squares

Utilizing multiple regression concepts, we explain how the sums of squares, F ratios, and proportions of variance in Table 2 are obtained for SPSS Option 10 and SAS Type I sums of squares. In explaining the results we will utilize Equation 5 and Figure 5.

Equation 5. $PV^2_T = R^2_A + R^2_{B.A} + R^2_{AXB.A,B} + (1 - R^2_{MAX})$

Referring to Table 2 under the SPSS Option 10/SAS Type I sums of squares, we find that the proportion of variance allocated to Factor A is found by dividing the sum of squares for Factor A (SS_A) by the total sum of squares (SS_{TOTAL}). This value includes the unique contribution of Factor A plus the proportion of the dependent variable variance accounted for by Factor A and the A x B interaction and the proportion of dependent variable variance accounted for by both Factors A and B. $R^2_A = SS_A / SS_{TOTAL} = 7.911/38.124 = .210$ with an F ratio = 5.006 and a probability of .012. In Figure 5, R^2_A = .210 represents Factor A plus the overlap area labeled 1, which is the proportion of variance shared by Factors A and B, plus the overlap area labeled 3, which is the proportion of variance shared jointly by Factor A and the A x B interaction.

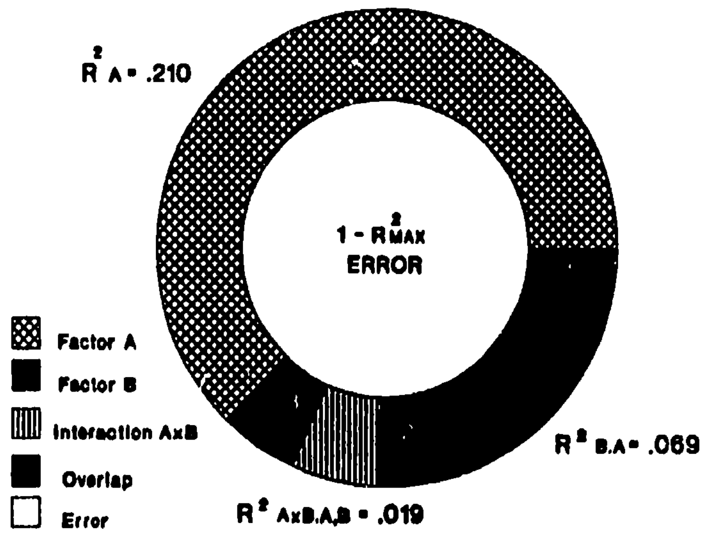

Figure 5. Variance partitioned using SPSS Option 10 and SAS Type I sums of squares.

Referring to Table 2, we find that the proportion of variance accounted for by Factor B, or the second variable entered on the procedure statement, is found by dividing the sum of squares for Factor B ($SS_{B.A}$) by the total sum of squares (SS_{TOTAL}). $R^2_{B.A} = SS_{B.A} / SS_{TOTAL} = 2.617/38.124 = .069$ with an F ratio = 3.312 and a probability of .078. When Factor B is the second variable entered into the equation Factor B variance is allocated as in the default option. In Figure 5, $R^2_{B.A}$ represents the proportion of the dependent variable variance accounted for uniquely by Factor B plus the area of overlap labeled 2, which is the proportion of the dependent variable variance accounted for jointly by Factor B and the A x B interaction. When using hierarchical approach, Option 10 from SPSS and Type I sums of squares from SAS, to represent Factor B or the second variable entered into the equation, the specific effects are the unique effects of Factor B while controlling for Factor A but not controlling for the A x B interaction. Numerically, $R^2_{B.A} = .069$.

In Table 2 the proportion of dependent variable variance accounted for uniquely by the A x B interaction utilizing SPSS Option 10 and SAS Type I sums of squares is identical to the results found with SPSS Option 9, SAS Type III sums of squares, and the default approach (SPSS). These explanations will not be repeated.

The area labeled error in Figure 5 refers to the proportion of variance in the dependent variable not allocated to either main effect or the A x B interaction. The proportion of dependent variable variance not explained and labeled as error is the same for all three methods and will not be repeated.

Discussion

When comparisons are made among the analytical methods in terms of proportions of variance accounted for by the effects, methodological differences become more apparent. With the default method of SPSS, $R^2_{A.B} = .179$ of the variance is allocated to Factor A. With SPSS Option 10 and SAS Type I sums of squares,

R^2_A = .210 of the variance is allocated to Factor A. When SPSS Option 9 and SAS Type III sums of squares are used, $R^2_{A.B,AXB}$ =.167 of the dependent variable variance is allocated to Factor A.

When F ratios and probabilities are evaluated in Table 2, we find that larger F ratios and smaller probabilities are assigned to Factor A with SPSS Option 10 and SAS Type I sums of squares. Smaller F ratios and larger probabilities are assigned to Factor A using SPSS Option 9 and SAS Type III sums of squares. The direction of these results are typical but are likely to be greater as the discrepancy in cell sizes increases.

When the variance accounted for by Factor B, or the second variable entered, the results are not as discrepant as they are for Factor A. The proportions of dependent variable variance allocated to Factor B using SPSS Option 10 and SAS Type I sums of squares are identical to the default SPSS ANOVA results, $R^2_{B.A}$ = .069, and are higher than the results from SPSS Option 9 ANOVA and SAS Type III sums of squares, $R^2_{B.A,AxB}$ = .063.

The interactions, which are evaluated after controlling for the main effects, are the same for both SAS and SPSS.

After some reflection upon the problem at hand and some practical experience with SPSS and SAS, it becomes fairly obvious that the order in which the independent variables are entered into the model can have a substantial impact on the proportions of dependent variable variance accounted for using SPSS Option 10 and SAS Type I sums of squares. The first variable entered into the model, Factor A in this case, has the opportunity to account for larger proportions of dependent variable variance than when the SPSS default option is used. This result is especially true when the cell sizes are grossly different.

Finally, different research questions are being asked of the data with each of the three methods. Utilizing the default option of the SPSS ANOVA procedure, researchers are answering the following questions:

1. For Factor A, what is the relationship between Factor A and the dependent variable after controlling for the main effect of Factor B but not controlling for the A x B interaction?

2. For Factor B, what is the relationship between Factor B and the dependent variable after controlling for the main effect of Factor A but not controlling for the A x B interaction?

3. What is the relationship between the A x B interaction and the dependent variable when controlling for the main effects of Factors A and B?

Using Option 9 of the SPSS ANOVA procedure and Type III sums of squares from the SAS GLM procedure, researchers are addressing the following research questions:

1. For Factor A, what is the relationship between Factor A and the dependent variable after controlling for the main effect of Factor B and the A x B interaction?

2. For Factor B, what is the relationship between Factor B and the dependent variable after controlling for the main effect of Factor A and the A x B interaction?

3. What is the relationship between the A x B interaction and the dependent variable when controlling for the main effects of Factors A and B?

Using Option 10 of the SPSS ANOVA procedure and Type I sums of squares from the SAS GLM procedure, researchers are addressing the following research questions:

1. For Factor A, what is the relationship between Factor A and the dependent variable?

2. For Factor B, what is the relationship between Factor B and the dependent variable after controlling for the main effect of Factor A?

3. What is the relationship between the A x B interaction and the dependent variable when controlling for the main effects of Factors A and B?

If researchers understand what research questions are being answered and exactly how the variance is being partitioned with analyses done with the ANOVA procedure of the Statistical Package for the Social Sciences and the GLM procedure of the Statistical Analysis System, they are much more likely to choose wisely among the options available to them.

References

Cochran, W. G., & Cox, G. M. (1950). Experimental designs. New York: Wiley.

Edington, E. S. (1974). A new tabulation of statistical procedures used in APA journals. American Psychologist, 29, 25-26.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14, 5-11.

Halpin G., & Halpin G. (1988, November). Evaluation of research and statistical methodologies. Paper presented at the meeting of the Mid-South Educational Research Association, Louisville, KY.

Lutz, J. G. (1979, February). The analysis of variance in non-orthogonal designs or which SPSS option should I use. Paper presented at the meeting of the Eastern Educational Research Association, Kiawah Island, SC.

Norusis M. J. (1990). SPSS/PC+ Statistics 4.0. Chicago: SPSS.

Norusis M. J. (1988). SPSS-X user's guide (3rd ed.). Chicago: SPSS.

Overall, J. E., & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. Psychological Bulletin, 72, 311-322.

SAS user's guide: Statistics (5th ed.). (1985). Cary, NC: SAS Institute.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. Educational Researcher, 9, 5-10.

Table 1

Scores Utilized in the Two-Factor Nonorthogonal ANOVA with SPSS and SAS

Factor A					
Level 1		Level 2		Level 3	
Factor B		Factor B		Factor B	
Level 1	Level 2	Level 1	Level 2	Level 1	Level 2
9.5	10.4	8.4	10.4	8.6	10.0
8.7	11.6	10.5	9.4	7.3	9.5
10.4	9.3	9.8	10.6	10.2	8.9
10.1	8.5	10.6	11.0	9.5	9.9
	10.3	11.4	11.1	9.8	10.6
	10.2	10.6	10.3	8.9	10.4
		10.4	10.6	9.7	
			10.7	10.0	
				7.1	

Table 2

Two-Factor ANOVA in Nonorthogonal Design Using SPSS and SAS

Source	SS	df	MS	F	p	R^2
Method 2: SPSS Default Option						
Factor A	6.806	2	3.403	4.307	.022	.179
Factor B	2.617	1	2.617	3.312	.078	.069
Interaction	.731	2	.366	.463	.634	.019
Error	26.866	34	.790			.705
Total	38.124	39				
Method 1: SPSS Option 9: SAS Type III SS						
Factor A	6.371	2	3.186	4.032	.027	.167
Factor B	2.391	1	2.391	3.026	.091	.063
Interaction	.731	2	.366	.463	.634	.019
Error	26.866	34	.790			.705
Total	38.124	39				
Method 3: SPSS Option 10: SAS Type I SS						
Factor A	7.911	2	3.955	5.006	.012	.210
Factor B	2.617	1	2.617	3.312	.078	.069
Interaction	.731	2	.366	.463	.643	.019
Error	26.866	34	.790			.705
Total	38.124	39				